**Weekly Report**

**2019.0102-2019.0106**

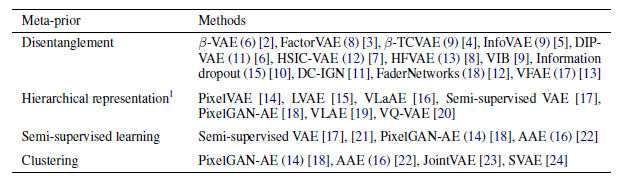
1. **This Week**

**AI FOR VIS Project**

1. 与王老师、宋老师、俊华、闽峰、浩哲讨论问题，梳理思路。
2. 问题思路如下：

图片中模式的定义需要domain knowledge，且这些模式可能发生在任意的时刻、设备上、设备之间，且具有多样性，所以人工去标注数据不切实际。我们需要用无监督的方法，去提取出这些图片中的模式，最后实现对这些图片中模式的挖掘。

1. 可行的解决方案：
   1. 第一步，通过传统机器学习方法中特征选择的这个部分，从可视化结果图片中定义特征（就是模式），例如hog sift等；第二步，需要人交互的检验、过滤这些特征（因子分析+人工找合适的因子并赋予每个因子意义）；第三步，基于这些特征做无监督聚类，约等于实现label；第四步，最后再通过神经网络在label后的数据集上做分类。整个过程迭代进行，思路与论文Deep Clustering for Unsupervised Learning of Visual Features相似。
   2. 方法3.1中，使用传统机器学习提取特征的方法，这一步可以换成深度学习手法提取特征，表达性是否会更强。可用的方法参考论文Unsupervised Feature Learning via Non-Parametric Instance Discrimination及其related work部分，包括了很多的self-supervised与unsupervised方法。
   3. 方法3.1中的第二步提到的因子分析方法，即论文Recent Advances in Autoencoder-Based Representation Learning中的disentanglement部分，如下图。（冯浩哲手上有现成的beta-vae代码，我请他帮忙先用beta-vae试做一下看效果，预计下周三1.9号出结果。）



* 1. 我认为方法3.1的缺点在于操作流程复杂，过程冗长。如果要结果过程复杂这个问题，可以去掉第四步，即做到聚类就停止，但去掉第四步的会造成，我们只能在一个图片中找到一种模式（因为一个图片只能有一个聚类类别），但加上第四步就可以做成multi-label分类问题，即可在一个图片中找到许多模式。

**潮流项目**

1.后端搭建完成，正在接上电科院PSASP软件中的计算程序（有提取出的exe应用程序）。

2.前后端完成接口沟通，下周完成前后端的数据请求。

3.后端主要是本科生王立冬在实现，近考试周进度稍微有些慢，我会适当提醒。

**南方电网项目**

1.数据、构思问题方向。

2.治军在整理电力数据可视化论文，并构思综述框架。

**其他**

1.准备组会报告。

**Working Hour: (except nap and eat time)**

9 小时/天 （周三至周五）

8 小时/周六

4小时/周日

总工作时长：39小时。

**论文阅读**

**1.EnsembleLens: Ensemble-based Visual Exploration of Anomaly Detection Algorithms with Multidimensional Data（组会报告论文）**

**2.Visual pattern mining in histology image collections using bag of features**

**Angel（使用传统机器学习方法提取图片特征）**

**3.ReVision: Automated Classification, Analysis and Redesign of Chart Images（06年李飞飞与jeffrey heer合作的可视化图表数据的分类与重新设计，最后实现一个集成的图表库，可以对效果不理想的可视化图表做重新设计）**

**4.Reverse-Engineering Visualizations: Recovering Visual Encodings from Chart Images**

**5.Discriminative k-shot learning using probabilistic models（jeffrey heer17年做的从可视化图标中提取视觉编码，提取出的视觉编码是code的形式，用的方法是cnn）。**

**6.** **Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery（如何使用生成模型实现特征提取，可用的生成模型包括gan、ae、vae）。**

**7.** **Unsupervised Feature Learning via Non-Parametric Instance Discrimination（如何用无监督的深度学习方法做特征提取）**

**8.** **Unsupervised Deep Embedding for Clustering Analysis（无监督的深度学习聚类方法）**

1. **Progress**

|  |  |  |
| --- | --- | --- |
| Task | Progress | Time |
| AI For Vis | 1.确定方法的基本思路 | 0226 |
| Power Flow Project | 1.前后端交接 | In one weeks |
| Southern Power Grid Project | 1.解读数据、构思问题  2.构思综述框架 | - |
| SQC paper | Delayed | - |